Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US04/021450

International filing date: 02 July 2004 (02.07.2004)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/485,052
Filing date: 03 July 2003 (03.07.2003)

Date of receipt at the International Bureau: 06 September 2004 (06.09.2004)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

1216336

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

August 24, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/485,052*
FILING DATE: *July 03, 2003*
RELATED PCT APPLICATION NUMBER: *PCT/US04/21450*

Certified by

Jon W Dudas

Acting Under Secretary of Commerce
for Intellectual Property
and Acting Director of the U.S.
Patent and Trademark Office

SUBSTITUTE PTO/SB/16 (5-03)

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR §1.53(c).

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Xaing-Dong	Fu	San Diego, CA

Additional inventors are being named on the 0 separately numbered sheets attached hereto

TITLE OF THE INVENTION (500 characters max)

Genome Mapping of Functional DNA Elements and Cellular Proteins

CORRESPONDENCE ADDRESS

Direct all correspondence to:

[X] Customer Number: 20985

OR

[ ] Firm *or* Individual Name					
Address					
Address					
City		State		ZIP	
Country	United States	Telephone		Fax	

ENCLOSED APPLICATION PARTS *(check all that apply)*

[X] Specification	*Number of Pages*	27	[] CD(s), Number		
[] Drawing(s)	*Number of Sheets*	5	[] Other (specify)	Appendix 25 pages	
[] Application Data Sheet. See 37 CFR 1.76.					

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

[X] Applicant Claims small entity status. See 37 CFR 1.27.

[X] A check or money order is enclosed to cover the filing fees.

[] The Director is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: 06-1050

[] Payment by credit card. Form PTO-2038 is attached.

FILING FEE AMOUNT ($): $80

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

[ ] No.

[X] Yes, the name of the U.S. Government agency and the Government contract number are: NIH/NCI 5R33CA88351

Respectfully submitted,

Signature _____ Date July 3, 2003

Typed Name Joseph R. Baker, Jr., Reg. No. 40,900

Telephone No. (858) 678-5070

Docket No. 15670-052P01

10310000.doc

CERTIFICATE OF MAILING BY EXPRESS MAIL

Express Mail Label No. EV348190275US

Date of Deposit 7/3/2003

Genome Mapping of Functional DNA Elements and Cellular Proteins

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

The invention was funded in part by Grant No. 5R33CA88351 awarded by the National Institutes of Health. The government may have certain rights in the invention.

TECHNICAL FIELD

This invention relates to mapping of proteins and DNA element in a genome.

BACKGROUND

Transcriptional regulation involves a large number of proteins or protein complexes specifically assembled at a given promoter to activate or suppress RNA synthesis. In a specific tissue or cell type, a promoter can be turned on by a sequence of specific recognition events. Transcription factors bind cis-acting regulatory sequences; these DNA binding proteins then recruit co-activator complexes and these pre-activation complexes then recruit the core transcription machinery. Such a sequential recruitment mechanism was demonstrated on the HO gene promoter during the cell cycle in yeast (Cosma et al., 1999). Similarly, a gene can be turned off by the recruitment of transcription co-repressor complexes through sequence-specific DNA binding proteins during repression involved chromatin remodeling factors that modify histones and a long term molecular memory may be established by epigenetic modification of a specific chromatin region(s) via DNA methylation. An advance in achieving progress in the area is the chromatin immunoprecipitation (ChIP) assay. This technology enables mapping of functional DNA elements that are engaging in interactions with specific DNA binding proteins and their associated protein complexes in vivo and has been applied to many individual case studies. In principle, this approach could lend itself to high-throughput detection methods, which would open up new opportunities for systems-level approaches to gene regulatory networks.

Researcher are seeking to identify various functional DNA elements embedded in the human genome, whether or not they are involved in gene expression, DNA replication, or establishment of chromosome territories in the cell. The method ideally suited for achieving the goal is the so-called ChIP-on-Chip technology, which is the ChIP assay coupled with high throughput detection on chips containing a microarray of human promoters.

The ChIP assay has been widely used in localizing in vivo binding sites for transcription factors. Briefly, cultured cells are treated with formaldehyde to induce crosslinking between

DNA and bound proteins in vivo. Treated cells are disrupted and nucleoproteins are recovered. Sonication is then used to randomly shear DNA into ~0.5 kb pieces. Because of covalent linkage induced by crosslinking, specific proteins remain associated with fragmented DNA. Specific antibodies against target proteins are used to immunoprecipitate DNA-protein complexes. Both starting and immunoprecipitated materials are analyzed by PCR using primers specific for a given DNA region(s) under investigation. A specific in vivo interaction can be inferred if immunoprecipitation results in a significant enrichment of the DNA fragment(s) in question.

The ChIP assay has been used to detect specific targets for transcription and DNA replication factors, chromatin remodeling factors, modified histones, methylated DNA, and the like. Furthermore, the assay has also been used to detect specific association of RNA binding proteins with DNA elements bridged by transcribing RNA because transcription and splicing are known to be spatially and temporarily coupled in the cell (Lei et al., 2001).

The ChIP-on-Chip technology has been used to address detailed mechanistic question on selected DNA target(s). However, starting and immunoprecipitated materials have to be analyzed by PCR one at a time, which requires the selection of a target set based on available functional information. In order to conduct unbiased search for specific interactions of a given transcription factor a ChIP-on-Chip system was developed. Briefly, using information from sequenced and annotated yeast genomes, individual intragenic sequences were PCR-amplified and spotted on glass to form a promoter microarray. Immunoprecipitated DNA fragments were linked by ligation with a primer-landing site on both ends, thereby permitting signal amplification by PCR (i.e., ligation-mediated PCR or LM-PCR). PCR amplified started and immunprecipitated materials were finally labeled with different fluorescence dye by random priming. Pooled PCR products were then hybridized to the promoter array to detect which promoters were specifically enriched by chromatin immunoprecipitation.

The ChIP-on-Chip technology requires 10^8 cells in each experiment, thus precluding analysis of development, tumorgenesis and stem cells where starting materials may be limited. In addition, microarray-based approaches will face the specificity issue.

A procedure referred to as RASL (for RNA Annealing Selection and Ligation) has been employed to address the specification issue generally associated with microarray approaches. In a 5' alternative splicing event, for example, there are two 5' splice sites in competition with a common 3' splice site. Three oligos are used to target to 20 nucleotide exonic sequences at each

splice site junction as diagrammed. In order to distinguish between the two competing 3' splice sites, a unique 20 nucleotide index sequence to each 5' oligo (1 or 2, labeled with red and green, respectively). The RASL assay includes the following processes: (1) Annealing, (2) Solid phase selection, (3) Ligation, (4) PCR amplification, and (5) Detection on a universal index array.

SUMMARY

The invention provides a method of detecting a polynucleotide-polypeptide interaction domain in a genome of an organism, comprising a) immunoprecipitating polynucleotides linked to a polypeptide; b) dissassociating the polynucleotide and polypeptide; c) contacting the polynucleotide with a primer pair under conditions whereby primer pair hybridize to the polynucleotide to form a first hybridization complex, each primer comprising at least two portions, a first portion comprising a target-specific oligonucleotide that is capable of hybridizing to a target polynucleotide, and a second portion comprising a universal primer landing site, the two primers are designed to be specific for an upstream and downstream segment of a target polynucleotide, one primer of the pair of primers comprising a first universal primer landing site and the second primer comprising a second universal primer landing site, wherein the universal landing sites are not the same, d) contacting the first hybridization complex with a ligase under conditions whereby primer pairs hybridized to the polynucleotide are ligated to for a ligated probe; e) amplifying the ligated probe with universal primers to generated an amplified-labeled product; f) contacting the amplified-labeled product with an array oligonucleotides to form assay complexes; and g) detecting said assay complexes as an indication, wherein the presence of complexes is indicative of DNA that binds the immunoprecipitated polypeptide.

The invention also provides a method of identifying a region of a genome of a living cell to which a polypeptide of interest binds, comprising the steps of: a) crosslinking DNA binding protein in the living cell to genomic DNA of the living cell, thereby producing DNA binding polypeptide crosslinked to genomic DNA; b) generating DNA fragments of the genomic DNA crosslinked to DNA binding polypeptide in a), thereby producing a mixture comprising DNA fragments to which DNA binding protein is bound; c) removing a DNA fragment to which the polypeptide of interest is bound from the mixture produced in b); d) separating the DNA fragment identified in c) from the polypeptide of interest; e) contacting the DNA with a primer pair under conditions whereby primer pair hybridize to the DNA to form a first hybridization

complex, each primer comprising at least two portions, a first portion comprising a target-specific oligonucleotide that is capable of hybridizing to a target polynucleotide, and a second portion comprising a universal primer landing site, the two primers are designed to be specific for an upstream and downstream segment of a target polynucleotide, one primer of the pair of primers comprising a first universal primer landing site and the second primer comprising a second universal primer landing site, wherein the universal landing sites are not the same; f) contacting the first hybridization complex with a ligase under conditions whereby primer pairs hybridized to the polynucleotide are ligated to for a ligated probe; g) amplifying the ligated probe of f); h) combining the amplified product of g) with DNA comprising a sequence complementary to genomic DNA of the cell, under conditions in which hybridization between the amplified product and a region of the sequence complementary to genomic DNA occurs to form a second hybridization complex; and i) identifying the second hybridization complex of h), wherein the second hybridization complex comprises the region of the genome in the cell to which the polypeptide of interest binds.

The invention further provides a method of identifying a region of a genome of a living cell to which a polypeptide of interest binds, comprising: a) crosslinking DNA binding polypeptides in the living cell to genomic DNA of the living cell, thereby producing DNA binding polypeptides crosslinked to genomic DNA; b) generating DNA fragments of the genomic DNA crosslinked to DNA binding polypeptides, thereby producing DNA fragments to which DNA binding polypeptides are bound; c) immunoprecipitating the DNA fragment produced using an antibody that specifically binds the polypeptide of interest; d) separating the DNA fragment identified in c) from the polypeptide of interest; e) contacting the DNA with a primer pair under conditions whereby primer pair hybridize to the DNA to form a first hybridization complex, each primer comprising at least two portions, a first portion comprising a target-specific oligonucleotide that is capable of hybridizing to a target polynucleotide, and a second portion comprising a universal primer landing site, the two primers are designed to be specific for an upstream and downstream segment of a target polynucleotide, one primer of the pair of primers comprising a first universal primer landing site and the second primer comprising a second universal primer landing site, wherein the universal landing sites are not the same; f) contacting the first hybridization complex with a ligase under conditions whereby primer pairs hybridized to the polynucleotide are ligated to for a ligated probe; g) amplifying the ligated

probe of f) using universal primers labeled with a detectable label; h) combining the amplified product of g) with DNA comprising a sequence complementary to genomic DNA of the cell, under conditions in which hybridization between the amplified product and a region of the sequence complementary to genomic DNA occurs to form a second hybridization complex; i) identifying the second hybridization complex of h) using methods specific for the label, wherein the second hybridization complex comprises the region of the genome in the cell to which the polypeptide of interest binds; and j) comparing the label intensity/amount measured in i) to the amount/intensity of a control, wherein amount/intensity of the label in a region of the genome which is greater than the amount/intensity of label of the control in the region indicates the region of the genome in the cell to which the polypeptide of interest binds.

The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

FIG. 1 depicts a general process of the invention.

FIG. 2-4 show results from a method of the invention.

DETAILED DESCRIPTION

The invention utilizes RASL technology in combination with ChIP-on-Chip technology. This combination is referred to herein as "ChIP-DASL".

Understanding how DNA-binding proteins control global gene expression, chromosomal replication and cellular proliferation would be facilitated by identification of the chromosomal locations at which these proteins function in vivo. Described herein is a genome-wide mapping method for regulated DNA elements and protein regulators.

The invention provides methods of examining the binding of proteins to DNA across a genome (e.g., the entire genome or a portion thereof, such as one or more chromosomes or a chromosome regions). In particular, the invention relates to a method of identifying a regulatory region (e.g., a protein or enhancer region) of genomic DNA to which a protein of interest binds. In one aspect, the invention looks at tissue related regulation. In another aspect, the invention looks at developmental related regulation. In yet another aspect, the invention looks at regulation of expression in a particular disease state or disorder.

Typically, proteins, which bind DNA are crosslinked to the cellular DNA. The resulting mixture will comprise both protein bound DNA and DNA that is not bound by protein. The mixture is then treated such as by shearing to generate smaller genomic fragments. As a result, DNA fragments crosslinked to DNA binding protein are generated and the DNA fragment(s) can be removed from the mixture. The resulting crosslinked fragments are then treated to separate the DNA binding proteins from the DNA. The DNA fragment is then combined with oligonucleotides primers comprising a sequence complementary to the DNA fragment under conditions in which hybridization between the DNA fragments and the oligonucleotide primer occurs.

The methods of the invention also provide the ability to determine whether a DNA binding protein is a transcription factor. The region of the sequence complementary to genomic DNA to which the DNA fragments hybridizes is identified wherein if the region of the genome is a regulatory region, then the protein of interest is a transcription factor.

The methods of the invention can be used to examine and/or identify DNA binding proteins across the entire genome of a eukaryotic organism. A variety of DNA binding proteins which bind to DNA can be analyzed. For example, any protein involved in DNA replication or transcription regulation can be examined in the methods of the invention.

Methods of crosslinking protein to DNA are known in the art. Such methods include, for example, UV light and formaldehyde.

Methods of separating/selecting DNA crosslinked to proteins are known in the art. For example, immunoprecipitation using an antibody (e.g., polyclonal, monoclonal) or antigen binding fragment thereof which binds (specifically) to a binding protein of interest, can be used. In addition, the protein of interest can be labeled or tagged using, for example, an antibody epitope (e.g., hemagglutinin (HA)).

In other methods for identification and isolation of regulatory regions, enrichment of regulatory DNA sequences takes advantage of the fact that the chromatin of actively-transcribed genes generally comprises acetylated histones. See, for example, Wolffe et al. (1996) Cell 84:817-819. In particular, acetylated H3 and H4 are enriched in the chromatin of transcribed genes, and chromatin comprising regulatory sequences is selectively enriched in acetylated H3. Accordingly, chromatin immunoprecipitation using antibodies to acetylated histones, particularly acetylated H3, can be used to obtain collections of sequences enriched in regulatory DNA.

Such methods generally involve fragmenting chromatin and then contacting the fragments with an antibody that specifically recognizes and binds to acetylated histones, particularly H3. The polynucleotides from the immunoprecipitate can subsequently be collected from the immunoprecipitate. Prior to fragmenting the chromatin, one can optionally crosslink the acetylated histones to adjacent DNA. Crosslinking of histones to the DNA within the chromatin can be accomplished according to various methods. One approach is to expose the chromatin to ultraviolet irradiation. Gilmour et aL (1984) Proc. Natil. Acad. Sci. USA 81:4275-4279. Other approaches utilize chemical crosslinking agents. Suitable chemical crosslinking agents include, but are not limited to, formaldehyde and psoralen. Solomon et al. (1985) Proc. NatL. Acad. Sci. USA 82:6470-6474; Solomon et al. (1988) Cell 53:937-947.

Fragmentation can be accomplished using established methods for fragmenting chromatin, including, for example, sonication, shearing and/or the use of restriction enzymes. The resulting fragments can vary in size, but using certain sonication techniques, fragments of approximately 200-400 nucleotide pairs are obtained.

Antibodies that can be used in the methods are commercially available from various sources. Examples of such antibodies include, but are not limited to, Anti Acetylated Histone H3, available from Upstate Biotechnology, Lake Placid, N.Y.

Polynucleotides in the methods described herein can be amplified using, for example, ligation-mediated polymerase chain reaction (e.g., see Current Protocols in Molecular Biology, Ausubel, F. M. et al., eds. 1991, the teachings of which are incorporated herein by reference).

Complementary polynucleotides (e.g. DNA) to that of the isolated DNA fragment to which the protein of interest binds can be hybridized using a variety of methods. For example, the complementary molecule can be immobilized on a glass slide (e.g., Corning Microarray Technology (CMT™) GAPS™) or on a microchip. Conditions of hybridization will typically include, for example, high stringency conditions and/or moderate stringency conditions. See e.g., pages 2.10.1-2.10.16 (see particularly 2.10.8-11) and pages 6.3.1-6 in Current Protocols in Molecular Biology). Factors such as probe length, base composition, percent mismatch between the hybridizing sequences, temperature and ionic strength influence the stability of hybridization. Thus, high or moderate stringency conditions can be determined empirically, and depend in part upon the characteristics of the polynucleotide (DNA, RNA) and the other nucleic acids to be assessed for hybridization.

In one aspect of the invention Chromatin Immunoprecipitation (ChIP) is used to obtain DNA fragments bound to proteins. Chromatin immunoprecipitation allows the detection of proteins that are bound to a particular region of DNA. It typically involves the following four steps: (1) formaldehyde cross-linking proteins to DNA in living cells, (2) disrupting and then sonicating the cells to yield small fragments of cross-linked DNA, (3) immunoprecipitating the protein-DNA crosslinks using an antibody which specifically binds the protein of interest, and (4) reversing the crosslinks.

Prior to or after ChIP the DNA is biotinylated. Once the DNA comprising polypeptides is removed from DNA that does not contain polypeptides, the biotinylated DNA is bound to solid surface with through biotin-streptavidin interactions.

The method combines a modified Chromatin Immunoprecipitation (ChIP) procedure, which has been previously used to study in vivo protein-DNA interactions at one or a small number of specific DNA sites, with DNA microarray analysis. Briefly, cells are fixed with formaldehyde, harvested by sonication, and DNA fragments that are crosslinked to a protein of interest are enriched by immunoprecipitation with a specific antibody. After reversal of the crosslinking, the enriched DNA is contacted with a primer pair under conditions whereby primer pair hybridize to the polynucleotide to form a first hybridization complex, each primer comprising at least two portions, a first portion comprising a target-specific oligonucleotide that is capable of hybridizing to a target polynucleotide, and a second portion comprising a universal primer landing site, the two primers are designed to be specific for an upstream and downstream segment of a target polynucleotide, one primer of the pair of primers comprising a first universal primer landing site and the second primer comprising a second universal primer landing site, wherein the universal landing sites are not the same, contacting the first hybridization complex with a ligase under conditions whereby primer pairs hybridized to the polynucleotide are ligated to for a ligated probe; amplifying the ligated probe with universal primers to generated an amplified-labeled product. For example, the amplification can take place using a fluorescent dye and ligation-mediated PCR (LM-PCR). A sample of DNA that has not been enriched by immunoprecipitation is subjected to LM-PCR in the presence of a different fluorophore, and both IP-enriched and unenriched pools of labeled-DNA are hybridized to a single DNA microarray (as discussed further herein). The IP-enriched/unenriched ratio of fluorescence intensity obtained from three independent experiments can be used with a weighted average analysis method to

calculate the relative binding of the polypeptide of interest to each sequence represented on the array.

Identification of a binding site for a particular defined transcription factor in cellular chromatin is indicative of the presence of regulatory sequences. This can be accomplished, for example, using the technique of chromatin immunoprecipitation. Briefly, this technique involves the use of a specific antibody to immunoprecipitate chromatin complexes comprising the corresponding antigen (in this case, the transcription factor of interest), and examination of the nucleotide sequences present in the immunoprecipitate. Immunoprecipitation of a particular sequence by the antibody is indicative of interaction of the antigen with that sequence. See, for example, O'Neill et al. in Methods in Enzymology, Vol. 274, Academic Press, San Diego, 1999, pp. 189-197; Kuo et al. (1999) Method 19:425-433; and Current Protocols in Molecular Biology, F. M. Ausubel et al., eds., Current Protocols, Chapter 21, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1998 Supplement).

As with the other methods, polynucleotides isolated from an immunoprecipitate, as described herein, can be cloned to generate a library and/or sequenced, and the resulting sequences used to populate a database as described in greater detail infra. Sequences adjacent to those detected by this method are also likely to be regulatory sequences. These can be identified by mapping the isolated sequences on the genome sequence for the organism from which the chromatin sample was obtained, and optionally entered into one or more databases.

The invention can be generally described as follows. A plurality of probes (also referred to herein as "hybridization probes") comprise at least two portions: a first portion comprises a target-specific oligonucleotide that is capable of hybridizing to a target polynucleotide, and a second portion comprising a "universal primer landing site". Two different hybridization probes are designed to be specific for an upstream and downstream segment of a target polynucleotide. An upstream hybridization probe will comprise a first universal primer landing site and the downstream hybridization probe will comprise a second universal primer landing site. The first and second universal landing sites are not the same. Examples of universal primer landing sites include the T7 and T3 universal primer landing sites. In one aspect of the invention, the first universal primer landing site is a T7 primer landing site and the second universal primer landing site is a T3 primer landing site.

These hybridization probes are hybridized to the isolated DNA obtained by ChIP, from a sample, without prior amplification, to form hybridization complexes. The non-hybridized DNA and hybridization probes are then removed. This is accomplished by using a streptavidin support that can specifically retain all biotinylated DNA, including hybrid complexes. Once the unhybridized probes are removed, the hybrids are subjected to ligation. The ligated probes can then be simultaneously amplified using universal primers that will hybridize to the upstream and downstream universal priming sequences. The resulting amplicons, which can be directly or indirectly labeled, can then be detected on arrays. This allows the detection and quantification of the target polynucleotides.

As will be appreciated by those in the art, target polynucleotides can be obtained from samples including, but not limited to, bodily fluids (including, but not limited to, blood, urine, serum, lymph, saliva, anal and vaginal secretions, perspiration and semen, of virtually any organism, with mammalian samples being preferred and human samples being particularly preferred). The sample may comprise individual cells, including primary cells (including bacteria), and cell lines, including, but not limited to, tumor cells of all types (particularly melanoma, myeloid leukemia, carcinomas of the lung, breast, ovaries, colon, kidney, prostate, pancreas and testes), cardiomyocytes, endothelial cells, epithelial cells, lymphocytes (T-cell and B cell), mast cells, eosinophils, vascular intimal cells, hepatocytes, leukocytes including mononuclear leukocytes, stem cells such as haemopoetic, neural, skin, lung, kidney, liver and myocyte stem cells; osteoclasts, chondrocytes and other connective tissue cells, keratinocytes, melanocytes, liver cells, kidney cells, and adipocytes. Suitable cells also include known research cells, including, but not limited to, Jurkat T cells, NIH3T3 cells, CHO, Cos, 923, HeLa, WI-38, Weri-1, MG-63, etc. See the ATCC cell line catalog, hereby expressly incorporated by reference.

The invention provides compositions and methods for detecting the presence or absence of polynucleotides that are bound by proteins in a sample.

A target polynucleotide includes polymeric form of nucleotides at least 20 bases in length. By "isolated polynucleotide" is meant a polynucleotide that is not immediately contiguous with either of the coding sequences with which it is immediately contiguous (one on the 5' end and one on the 3' end) in the naturally occurring genome of the organism from which it is derived. The term therefore includes, for example, a recombinant DNA which is incorporated into a vector; into an automatically replicating plasmid or virus; or into the genomic DNA of a

prokaryote or eukaryote, which exists as a separate molecule (e.g., a cDNA) independent of other sequences, as well as genomic fragments that may be present in solution or on microarray chips. The nucleotides of the invention can be ribonucleotides, deoxyribonucleotides, or modified forms of either nucleotide. The term includes single and double stranded forms of DNA.

The term polynucleotide(s) generally refers to any polyribonucleotide or polydeoxyribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. Thus, for instance, polynucleotides as used herein refers to, among others, single-and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions.

In addition, polynucleotide is used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide.

As used herein, the term polynucleotide includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotides" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein.

It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including simple and complex cells, inter alia.

The target polynucleotide sequence may also be comprised of different target domains, that may be adjacent (i.e. contiguous) or separated. For example, in the OLA techniques outlined below, a first hybridization probe may hybridize to a first target domain and a second hybridization probe may hybridize to a second target domain; either the domains are adjacent, or they may be separated by one or more nucleotides, coupled with the use of a polymerase and

dNTPs, as is more fully outlined below. The terms "first" and "second" are not meant to confer an orientation of the sequences with respect to the 5'-3' orientation of the target polynucleotide. For example, assuming a 5'-3' orientation of the complementary target sequence, the first target domain may be located either 5' to the second domain, or 3' to the second domain. In addition, as will be appreciated by those in the art, the probes on the surface of the array may be attached in either orientation, either such that they have a free 3' end or a free 5' end; in some embodiments, the probes can be attached at one ore more internal positions, or at both ends.

The target polynucleotide is prepared using known techniques. For example, the sample may be treated to lyse the cells, using known lysis buffers, sonication, electroporation, etc., with purification and amplification as outlined below occurring as needed, as will be appreciated by those in the art. In addition, the reactions outlined herein may be accomplished in a variety of ways, as will be appreciated by those in the art. Components of the reaction may be added simultaneously, or sequentially, in any order, with typical embodiments outlined below. In addition, the reaction may include a variety of other reagents which may be included in the assays. These include reagents like salts, buffers, neutral proteins, e.g. albumin, detergents, etc., which may be used to facilitate optimal hybridization and detection, and/or reduce non-specific or background interactions. Also reagents that otherwise improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc., may be used, depending on the sample preparation methods and purity of the target.

In addition, in most embodiments, double stranded target polynucleotides are denatured to render them single stranded so as to permit hybridization of the primers and other probes of the invention. A typical embodiment utilizes a thermal step, generally by raising the temperature of the reaction to about 95°C, although pH changes and other techniques may also be used.

As will be appreciated by those in the art, all of these nucleic acid analogs may find use as probes in the invention. In addition, mixtures of naturally occurring nucleic acids and analogs can be made. Alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occurring nucleic acids and analogs may be made.

Peptide nucleic acids (PNA) which includes peptide nucleic acid analogs can be used. These backbones are substantially non-ionic under neutral conditions, in contrast to the highly charged phosphodiester backbone of naturally occurring nucleic acids. This results in two advantages. First, the PNA backbone exhibits improved hybridization kinetics. PNAs have larger

changes in the melting temperature (T_m) for mismatched versus perfectly matched basepairs. DNA and RNA typically exhibit a 2-4°C drop in T_m for an internal mismatch. With the non-ionic PNA backbone, the drop is closer to 7-9°C. Similarly, due to their non-ionic nature, hybridization of the bases attached to these backbones is relatively insensitive to salt concentration.

The hybridization probe may contain any combination of deoxyribo- and ribo-nucleotides, and any combination of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xathanine hypoxathanine, isocytosine, isoguanine, etc. In one embodiment, isocytosine and isoguanine are used in nucleic acids designed to be complementary to other probes, rather than target sequences, as this reduces non-specific hybridization, as is generally described in U.S. Pat. No. 5,681,702. As used herein, the term "nucleoside" includes nucleotides as well as nucleoside and nucleotide analogs, and modified nucleosides such as amino modified nucleosides. In addition, "nucleoside" includes non-naturally occurring analog structures. Thus for example the individual units of a peptide nucleic acid, each containing a base, are referred to herein as a nucleoside.

Probes and primers of the invention are designed to have at least a portion be complementary to a target polynucleotide, such that hybridization of the target polynucleotide and the probes of the invention occurs. As outlined below, this complementarity need not be perfect; there may be any number of base pair mismatches which will interfere with hybridization between the target polynucleotide and the single stranded hybridization probe of the invention. Thus, by "substantially complementary" herein is meant that the probes are sufficiently complementary to the target polynucleotide sequences to hybridize under normal reaction conditions, and preferably give the required specificity.

A variety of hybridization conditions may be used in the invention, including high, moderate and low stringency conditions; see for example Maniatis et al., Molecular Cloning: A Laboratory Manual, 2d Edition, 1989, and Short Protocols in Molecular Biology, ed. Ausubel, et al, hereby incorporated by reference. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting

point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target hybridize to the polyadenylated mRNA target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30° C for short probes (e.g. 10 to 50 nucleotides) and at least about 60° C for long probes (e.g. greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of helix destabilizing agents such as formamide. The hybridization conditions may also vary when a non-ionic backbone, i.e. PNA is used, as is known in the art. In addition, cross-linking agents may be added after target binding to cross-link, i.e. covalently attach, the two strands of the hybridization complex.

Thus, the assays are generally run under stringency conditions which allows formation of the first hybridization complex only in the presence of target. Stringency can be controlled by altering a step parameter that is a thermodynamic variable, including, but not limited to, temperature, formamide concentration, salt concentration, chaotropic salt concentration, pH, organic solvent concentration, etc.

These parameters may also be used to control non-specific binding, as is generally outlined in U.S. Pat. No. 5,681,697. Thus it may be desirable to perform certain steps at higher stringency conditions to reduce non-specific binding.

The size of the primer and probe nucleic acid may vary, as will be appreciated by those in the art with each portion of the probe and the total length of the probe in general varying from 5 to 500 nucleotides in length. Each portion is preferably between 10 and 100 being preferred, between 15 and 50 being particularly useful and from 10 to 35 being typically used depending on the use and amplification technique. Thus, for example, the universal priming sites of the probes are each preferably about 15-25 nucleotides in length, with 20 being used most frequently. The adapter sequences of the probes are preferably from 15-25 nucleotides in length, with 20 being most common. The target specific portion of the probe is typically from 15-50 nucleotides in length, with from 30 to 40 being most common.

Accordingly, the invention provides first hybridization probe sets. By "probe set" herein is meant a plurality of hybridization probes that are used in a particular multiplexed assay. In this

context, plurality means at least two, with more than 10 being typically, depending on the assay, sample and purpose of the test.

Accordingly, the invention provides first hybridization probe sets that comprise universal priming sites. By "universal priming site" herein is meant a sequence of the probe that will bind a PCR primer for amplification. Each probe preferably comprises an upstream universal priming site (UUP) and a downstream universal priming site (DUP). Again, "upstream" and "downstream" are not meant to convey a particular 5'-3' orientation, and will depend on the orientation of the system. Typically, only a single UUP sequence and a single DUP sequence is used in a probe set, although as will be appreciated by those in the art, different assays or different multiplexing analysis may utilize a plurality of universal priming sequences. In addition, the universal priming sites are typically located at the 5' and 3' termini of the hybridization probe (or the ligated probe), as only sequences flanked by priming sequences will be amplified.

In addition, universal priming sequences are generally chosen to be as unique as possible given the particular assays and host genomes to ensure specificity of the assay. In general, universal priming sequences range in size from about 5 to about 35 basepairs, with from about 15 to about 20 being particularly preferred.

As will be appreciated by those in the art, the orientation of the two priming sites is different. That is, one PCR primer will directly hybridize to the first priming site, while the other PCR primer will hybridize to the complement of the second priming site. Stated differently, the first priming site is in sense orientation, and the second priming site is in antisense orientation.

In addition to the universal priming sites, the hybridization probes comprise at least a first target-specific sequence. As outlined below, hybridization probes each comprise a target-specific sequence. As will be appreciated by those in the art, the target-specific sequence may take on a wide variety of formats, depending on the use of probe. For example through a primer selection program, a specific 40-mer DNA sequence can be selected to represent a given region (such as promoter) in the human genome. The process will verify its uniqueness by allowing at least 4 evenly distributed mismatches in related sequences in the genome after the BLAST search against the human genome database(s). Selected sequences also avoid small repeats, have a T_m in a defined range (e.g., between about 55 and 65 °C), and contain minimized secondary structure (calculated by ΔG). In parallel, amino-derived oligos will be synthesized and spotted

onto the Motorola 3D codelink slide to form an oligo-based promoter array. This 40-mer is essentially split in two to provide two 20-mer target specific sequences that are combined with universal primers and thus become the upstream and downstream hybridization probes.

The two hybridization probes can be used in OLA assay systems. The basic OLA method can be run at least two different ways; in a first embodiment, only one strand of a target sequence is used as a template for ligation; alternatively, both strands may be used; the latter is generally referred to as Ligation Chain Reaction or LCR. See generally U.S. Pat. Nos. 5,185,243 and 5,573,907; EP 0 320 308 B1; EP 0 336 731 B1; EP 0 439 182 B1; WO 90/01069; WO 89/12696; and WO 89/09835, all of which are incorporated by reference. The discussion below focuses on OLA, but as those in the art will appreciate, this can easily be applied to LCR as well.

In this embodiment, the hybridization probes comprise at least a first hybridization probe and a second hybridization probe. The method is based on the fact that two probes can be ligated together, if they are hybridized to a target polynucleotide and if perfect complementarity exists at the junction.

In one embodiment, the two hybridization probes are designed each with a target specific portion. The first hybridization probe is designed to be substantially complementary to a first target domain of a target polynucleotide, and the second hybridization probe is substantially complementary to a second target domain of a target polynucleotide. As outlined herein, in one embodiment the first and second target domains are directly adjacent, e.g. they have no intervening nucleotides. In an alternative embodiment, the first and second target domains are indirectly adjacent, e.g. there are intervening nucleotides, and the system includes a polymerase and dNTPs that can be used to "fill in" the gap prior to ligation.

In this embodiment, at least a first hybridization probe is hybridized to the first target domain and a second hybridization probe is hybridized to the second target domain. If perfect complementarity exists at the junction, a ligation structure is formed such that the two probes can be ligated together to form a ligated probe. If this complementarity does not exist, no ligation structure is formed and the probes are not ligated together to an appreciable degree. This may be done using heat cycling, to allow the ligated probe to be denatured off the target polynucleotide such that it may serve as a template for further reactions. In addition, as is more fully outlined below, this method may also be done using three hybridization probes or hybridization probes

that are separated by one or more nucleotides, if dNTPs and a polymerase are added (this is sometimes referred to as "Genetic Bit" analysis).

In general, each target specific sequence of a hybridization probe is at least about 5 nucleotides long, with sequences of about 15 to 30 being typical and 20 being especially common.

In another embodiment, the two hybridization probes are not directly adjacent. In this embodiment, they may be separated by one or more bases. The addition of dNTPs and a polymerase, as outlined below for the amplification reactions, followed by the ligation reaction, allows the formation of the ligated probe.

Once the non-hybridized probes (and additionally other sequences from the sample that are not of interest) are removed, the hybridization complexes are denatured and the ligated probes are amplified to form amplicons, which are then detected. This can be done in one of several ways, including PCR amplification and rolling circle amplification. In addition, as outlined below, labels can be incorporated into the amplicons in a variety of ways.

In one embodiment, the target amplification technique is PCR. The polymerase chain reaction (PCR) is widely used and described, and involves the use of primer extension combined with thermal cycling to amplify a target sequence; see U.S. Pat. Nos. 4,683,195 and 4,683,202, and PCR Essential Data, J. W. Wiley & sons, Ed. C. R. Newton, 1995, all of which are incorporated by reference.

In general; PCR may be briefly described as follows. The double stranded hybridization complex is denatured, generally by raising the temperature, and then cooled in the presence of an excess of a PCR primer, which then hybridizes to the first universal priming site. A DNA polymerase then acts to extend the primer with dNTPs, resulting in the synthesis of a new strand forming a hybridization complex. The sample is then heated again, to disassociate the hybridization complex, and the process is repeated. By using a second PCR primer for the complementary target strand that hybridizes to the second universal priming site, rapid and exponential amplification occurs. Thus PCR steps are denaturation, annealing and extension. The particulars of PCR are well known, and include the use of a thermostable polymerase such as Taq I polymerase and thermal cycling. Suitable DNA polymerases include, but are not limited to, the Klenow fragment of DNA polymerase I, SEQUENASE 1.0 and SEQUENASE 2.0 (U.S. Biochemical), T5 DNA polymerase and Phi29 DNA polymerase.

The reaction is initiated by introducing the ligated probe to a solution comprising the universal primers, a polymerase and a set of nucleotides. By "nucleotide" in this context herein is meant a deoxynucleoside-triphosphate (also called deoxynucleotides or dNTPs, e.g. dATP, dTTP, dCTP and dGTP). In some embodiments, as outlined below, one or more of the nucleotides may comprise a detectable label, which may be either a primary or a secondary label. In addition, the nucleotides may be nucleotide analogs, depending on the configuration of the system. Similarly, the primers may comprise a primary or secondary label.

Accordingly, the PCR reaction requires at least one and typically two PCR primers, a polymerase, and a set of dNTPs. As outlined herein, the primers may comprise the label, or one or more of the dNTPs may comprise a label.

These embodiments also have the advantage that unligated probes need not necessarily be removed, as in the absence of the target, no significant amplification will occur. These benefits may be maximized by the design of the probes; for example, in the first embodiment, when there is a single hybridization probe, placing the universal priming site close to the 5' end of the probe since this will only serve to generate short, truncated pieces in the absence of the ligation reaction.

Labeling of the amplicon can be accomplished in a variety of ways; for example, the polymerase may incorporate labelled nucleotides (dNTPs), or alternatively, the universal primer itself comprises a label.

The polymerase can be any polymerase, but typically will lack 3' exonuclease activity. Examples of suitable polymerase include but are not limited to exonuclease minus DNA Polymerase I large (Klenow) Fragment, Phi29 DNA polymerase, Taq DNA Polymerase and the like. In addition, in some embodiments, a polymerase that will replicate single-stranded DNA (i.e. without a primer forming a double stranded section) can be used.

By "label" or "detectable label" is meant a moiety that allows detection. This may be a primary label or a secondary label. Accordingly, detection labels may be primary labels (i.e. directly detectable) or secondary labels (indirectly detectable).

In one embodiment, the detection label is a primary label. A primary label is one that can be directly detected, such as a fluorophore. In general, labels fall into three classes: a) isotopic labels, which may be radioactive or heavy isotopes; b) magnetic, electrical, thermal labels; and c) colored or luminescent dyes. Labels can also include enzymes (horseradish peroxidase, etc.) and

magnetic particles. Typical labels include chromophores or phosphors but are typically fluorescent dyes. Suitable dyes for use in the invention include, but are not limited to, fluorescent lanthanide complexes, including those of Europium and Terbium, fluorescein, rhodamine, tetramethylrhodamine, eosin, erythrosin, coumarin, methyl-coumarins, quantum dots (also referred to as "nanocrystals": see U.S. Ser. No. 09/315,584, hereby incorporated by reference), pyrene, Malacite green, stilbene, Lucifer Yellow, Cascade Blue™, Texas Red, Cy dyes (Cy3, Cy5, etc.), alexa dyes, phycoerythin, bodipy, and others described in the 6th Edition of the Molecular Probes Handbook by Richard P. Haugland, hereby expressly incorporated by reference.

A secondary label is one that is indirectly detected; for example, a secondary label can bind or react with a primary label for detection, can act on an additional product to generate a primary label (e.g. enzymes), or may allow the separation of the compound comprising the secondary label from unlabeled materials, etc. Secondary labels include, but are not limited to, one of a binding partner pair such as biotin/streptavidin; chemically modifiable moieties; nuclease inhibitors, enzymes such as horseradish peroxidase, alkaline phosphatases, lucifierases, etc.

The secondary label is typically a binding partner pair. For example, the label may be a hapten or antigen, which will bind its binding partner. For example, suitable binding partner pairs include, but are not limited to: antigens (such as proteins (including peptides)) and antibodies (including fragments thereof (FAbs, etc.)); proteins and small molecules, including biotin/streptavidin; enzymes and substrates or inhibitors; other protein-protein interacting pairs; receptor-ligands; and carbohydrates and their binding partners. Nucleic acid--nucleic acid binding proteins pairs are also useful. In general, the smaller of the pair is attached to the NTP for incorporation into the primer. Typical binding partner pairs include, but are not limited to, biotin (or imino-biotin) and streptavidin, digeoxinin and Abs, and Prolinx™ reagents. For example, the binding partner pair comprises biotin or imino-biotin and a fluorescently labeled streptavidin. Imino-biotin disassociates from streptavidin in pH 4.0 buffer while biotin requires harsh denaturants (e.g. 6 M guanidinium HCl, pH 1.5 or 90% formamide at 95 °C.).

Labeling can occur in a variety of ways, as will be appreciated by those in the art. In general, labeling can occur in one of two ways: labels are incorporated into primers such that the

amplification reaction results in amplicons that comprise the labels or labels are attached to dNTPs and incorporated by the polymerase into the amplicons.

The amplified DNA can be fluorescently labeled by including fluorescently-tagged nucleotides in the LM-PCR reaction or by fluorescently labeling the universal primers. Finally, the labeled DNA was hybridized to a DNA microarray containing spots representing all or a subset (e.g., a chromosome or chromosomes) of the genome. The fluorescent intensity of each spot on the microarray relative to a non-immunoprecipitated control demonstrated whether the protein of interest bound to the DNA region located at that particular spot. Hence, the methods described herein allow the detection of protein-DNA interactions across the entire genome.

The invention provides methods and compositions useful in the detection of polynucleotides that interact with polypeptide molecules. The process comprises immunopreciptiating DNA that is crosslinked to a polypeptide; dissociating the polypeptide from the DNA; hybridizing a pair of probes each comprising, for example, a 20-mer target sequence and a universal primer to the DNA; ligating the probes to form ligated probes; amplifying the ligated probes using the universal primers comprising a label; and contacting a DNA microarray with the amplified-labeled product. The amplified products are attached (via hybridization) to an array site comprising substantially complementary DNA sequence to those of the hybridization probe target sequence.

Accordingly, the invention provides array compositions comprising at least a first substrate with a surface comprising individual sites. By "array" or "biochip" herein is meant a plurality of polynucleotides or oligonucleotide in an array format; the size of the array will depend on the composition and end use of the array. Nucleic acids arrays are known in the art, and can be classified in a number of ways; both ordered arrays (e.g. the ability to resolve chemistries at discrete sites), and random arrays are included. Ordered arrays include, but are not limited to, those made using photolithography techniques (Affymetrix GeneChip™), spotting techniques (Synteni and others), printing techniques (Hewlett Packard and Rosetta), three dimensional "gel pad" arrays, etc. In addition, liquid arrays find use in the invention.

Generally, the array will comprise from two to as many as a billion or more different sequences, depending on the size of the substrate, as well as the end use of the array, thus very high density, high density, moderate density, low density and very low density arrays may be used. For example, very high density arrays are from about 10,000,000 to about 2,000,000,000,

with from about 100,000,000 to about 1,000,000,000 being typical (all numbers being in square cm). High density arrays range about 100,000 to about 10,000,000, with about 1,000,000 to about 5,000,000 being typical. Moderate density arrays range from about 10,000 to about 100,000 being typical, and from about 20,000 to about 50,000 being most common. Low density arrays are generally less than 10,000, with from about 1,000 to about 5,000 being typical. Very low density arrays are less than 1,000, with from about 10 to about 1000 being typical, and from about 100 to about 500 being most common.

By "substrate" or "solid support" is meant any material that can be modified to contain discrete individual sites appropriate for the attachment or association of oligonucleotides, polynucleotides, or other organic polymers and is amenable to at least one detection method. Possible substrates include, but are not limited to, glass and modified or functionalized glass, plastics (including acrylics, polystyrene and copolymers of styrene and other materials, polypropylene, polyethylene, polybutylene, polyurethanes, Teflon, etc.), polysaccharides, nylon or nitrocellulose, resins, silica or silica-based materials including silicon and modified silicon, carbon, metals, inorganic glasses, plastics, optical fiber bundles, and a variety of other polymers. In general, the substrates allow optical detection and do not themselves appreciably fluoresce.

Generally the substrate is flat (planar), although as will be appreciated by those in the art, other configurations of substrates may be used as well; for example, three dimensional configurations can be used, for example by embedding the beads in a porous block of plastic that allows sample access to the beads and using a confocal microscope for detection. Similarly, the beads may be placed on the inside surface of a tube, for flow-through sample analysis to minimize sample volume.

Generally, the array of array compositions can be configured in several ways; see for example U.S. Ser. No. 09/473,904, which is incorporated by reference. For example, a first substrate comprising a plurality of assay locations (sometimes also referred to herein as "assay wells"), such as a microtiter plate, is configured such that each assay location contains an individual array. That is, the assay location and the array location are the same. For example, the plastic material of the microtiter plate can be formed to contain a plurality of "wells" in the bottom of each of the assay wells.

In another aspect, the number of individual arrays is set by the size of the microtiter plate used; thus, 96 well, 384 well and 1536 well microtiter plates utilize composite arrays comprising

96, 384 and 1536 individual arrays, although as will be appreciated by those in the art, not each microtiter well need contain an individual array. It should be noted that the composite arrays can comprise individual arrays that are identical, similar or different. That is, in some embodiments, it may be desirable to do the same 2,000 assays on 96 different samples; alternatively, doing 192,000 experiments on the same sample (i.e. the same sample in each of the 96 wells) may be desirable. Alternatively, each row or column of the composite array could be the same, for redundancy/quality control. As will be appreciated by those in the art, there are a variety of ways to configure the system. In addition, the random nature of the arrays may mean that the same population of beads may be added to two different surfaces, resulting in substantially similar but perhaps not identical arrays.

In use the amplified-labeled product is exposed to the array comprising the substantially complementary polynucleotide/oligonucleotide sequence as in the hybridization probe(s). The product and polynucleotide/oligonucleotide in the microarray can hybridize (either directly or indirectly) resulting in a change in the optical signal of a particular microarray location.

A number of embodiments of the invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. Appendix A attached hereto further describes the invention and is considered a part of this disclosure and is incorporated herein in its entirety. Accordingly, other embodiments are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A method of detecting a polynucleotide-polypeptide interaction domain in a genome of an organism, comprising

a) immunoprecipitating polynucleotides linked to a polypeptide;

b) dissassociating the polynucleotide and polypeptide;

c) contacting the polynucleotide with a primer pair under conditions whereby primer pair hybridize to the polynucleotide to form a first hybridization complex, each primer comprising at least two portions, a first portion comprising a target-specific oligonucleotide that is capable of hybridizing to a target polynucleotide, and a second portion comprising a universal primer landing site, the two primers are designed to be specific for an upstream and downstream segment of a target polynucleotide, one primer of the pair of primers comprising a first universal primer landing site and the second primer comprising a second universal primer landing site, wherein the universal landing sites are not the same,

d) contacting the first hybridization complex with a ligase under conditions whereby primer pairs hybridized to the polynucleotide are ligated to for a ligated probe;

e) amplifying the ligated probe with universal primers to generated an amplified-labeled product;

f) contacting the amplified-labeled product with an array oligonucleotides to form assay complexes; and

g) detecting said assay complexes as an indication, wherein the presence of complexes is indicative of DNA that binds the immunoprecipitated polypeptide.

2. A method of identifying a region of a genome of a living cell to which a polypeptide of interest binds, comprising the steps of:

a) crosslinking DNA binding protein in the living cell to genomic DNA of the living cell, thereby producing DNA binding polypeptide crosslinked to genomic DNA;

b) generating DNA fragments of the genomic DNA crosslinked to DNA binding polypeptide in a), thereby producing a mixture comprising DNA fragments to which DNA binding protein is bound;

c) removing a DNA fragment to which the polypeptide of interest is bound from the mixture produced in b);

d) separating the DNA fragment identified in c) from the polypeptide of interest;

e) contacting the DNA with a primer pair under conditions whereby primer pair hybridize to the DNA to form a first hybridization complex, each primer comprising at least two portions, a first portion comprising a target-specific oligonucleotide that is capable of hybridizing to a target polynucleotide, and a second portion comprising a universal primer landing site, the two primers are designed to be specific for an upstream and downstream segment of a target polynucleotide, one primer of the pair of primers comprising a first universal primer landing site and the second primer comprising a second universal primer landing site, wherein the universal landing sites are not the same;

f) contacting the first hybridization complex with a ligase under conditions whereby primer pairs hybridized to the polynucleotide are ligated to for a ligated probe;

g) amplifying the ligated probe of f);

h) combining the amplified product of g) with DNA comprising a sequence complementary to genomic DNA of the cell, under conditions in which hybridization between the amplified product and a region of the sequence complementary to genomic DNA occurs to form a second hybridization complex; and

i) identifying the second hybridization complex of h), wherein the second hybridization complex comprises the region of the genome in the cell to which the polypeptide of interest binds.

3. The method of claim 2, wherein the cell is a eukaryotic cell.

4. The method of claim 2, wherein the polypeptide of interest is a transcription factor.

5. The method of claim 2, wherein the DNA binding polypeptide of the cell is crosslinked to the genome of the cell using formaldehyde.

6. The method of claim 1 or 2, wherein the DNA or polynucleotide to which the polypeptide is bound removed or separated using an antibody which binds to the polypeptide.

7. The method of claim 2, wherein the DNA fragment of g) is amplified using ligation-mediated polymerase chain reaction.

8. The method of claim 2, wherein the second hybridization complex is formed on a DNA microarray.

9. A method of identifying a region of a genome of a living cell to which a polypeptide of interest binds, comprising:

a) crosslinking DNA binding polypeptides in the living cell to genomic DNA of the living cell, thereby producing DNA binding polypeptides crosslinked to genomic DNA;

b) generating DNA fragments of the genomic DNA crosslinked to DNA binding polypeptides, thereby producing DNA fragments to which DNA binding polypeptides are bound;

c) immunoprecipitating the DNA fragment produced using an antibody that specifically binds the polypeptide of interest;

d) separating the DNA fragment identified in c) from the polypeptide of interest;

e) contacting the DNA with a primer pair under conditions whereby primer pair hybridize to the DNA to form a first hybridization complex, each primer comprising at least two portions, a first portion comprising a target-specific oligonucleotide that is capable of hybridizing to a target polynucleotide, and a second portion comprising a universal primer landing site, the two primers are designed to be specific for an upstream and downstream segment of a target polynucleotide, one primer of the pair of primers comprising a first universal primer landing site and the second primer comprising a second universal primer landing site, wherein the universal landing sites are not the same;

f) contacting the first hybridization complex with a ligase under conditions whereby primer pairs hybridized to the polynucleotide are ligated to for a ligated probe;

g) amplifying the ligated probe of f) using universal primers labeled with a detectable label;

h) combining the amplified product of g) with DNA comprising a sequence complementary to genomic DNA of the cell, under conditions in which hybridization between the amplified product and a region of the sequence complementary to genomic DNA occurs to form a second hybridization complex;

i) identifying the second hybridization complex of h) using methods specific for the label, wherein the second hybridization complex comprises the region of the genome in the cell to which the polypeptide of interest binds; and

j) comparing the label intensity/amount measured in i) to the amount/intensity of a control, wherein amount/intensity of the label in a region of the genome which is greater than the amount/intensity of label of the control in the region indicates the region of the genome in the cell to which the polypeptide of interest binds.

ABSTRACT

The invention provides methods of examining the binding of proteins to DNA across a genome (e.g., the entire genome or a portion thereof, such as one or more chromosomes or a chromosome regions). In particular, the invention relates to a method of identifying a regulatory region (e.g., a protein or enhancer region) of genomic DNA to which a protein of interest binds. In one aspect, the invention looks at tissue related regulation. In another aspect, the invention looks at developmental related regulation. In yet another aspect, the invention looks at regulation of expression in a particular disease state or disorder.

FIGURE 1A

The ChIP-DASL Technology

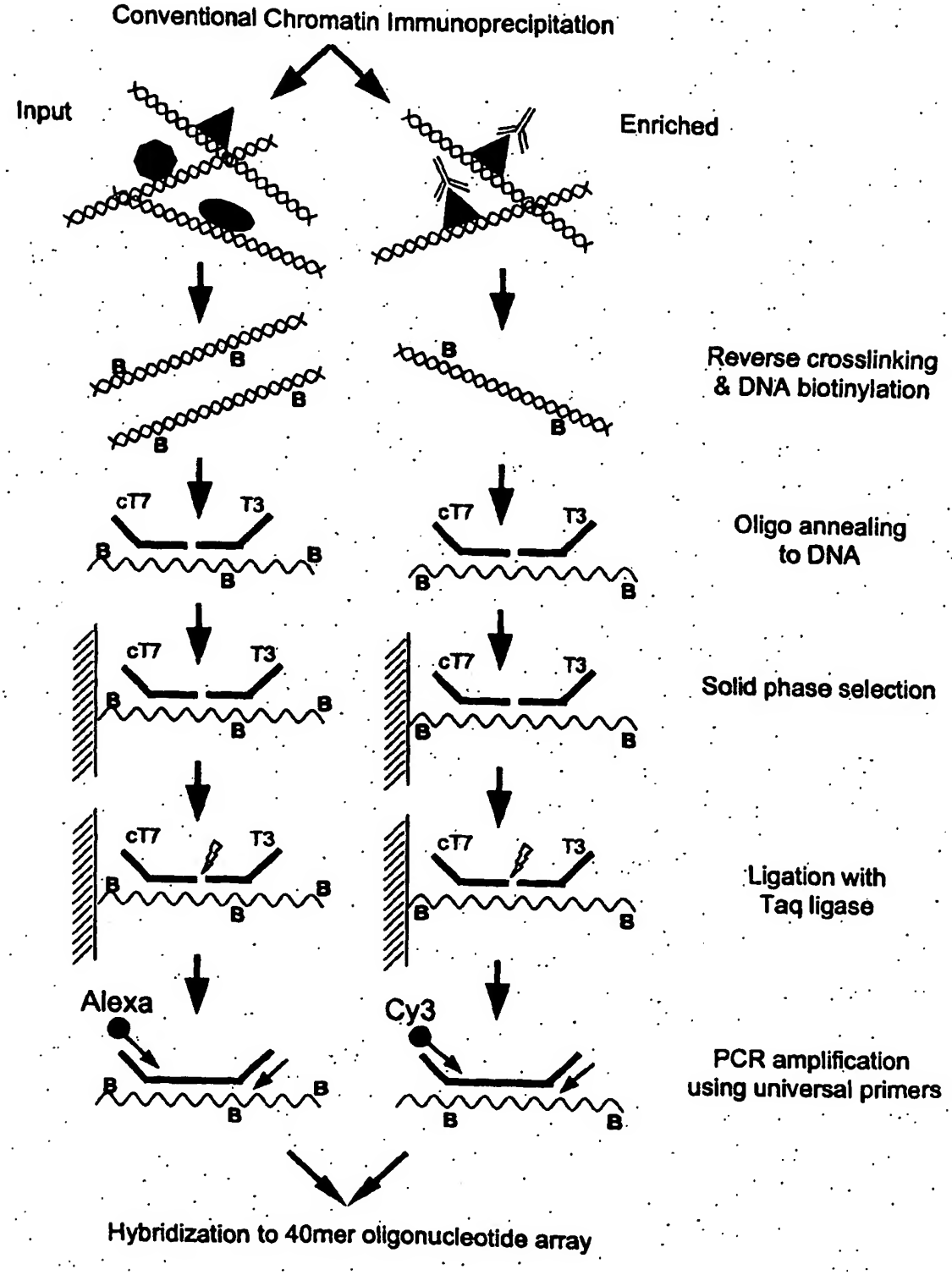


FIGURE 1B

Applications of the ChIP-DASL Technology

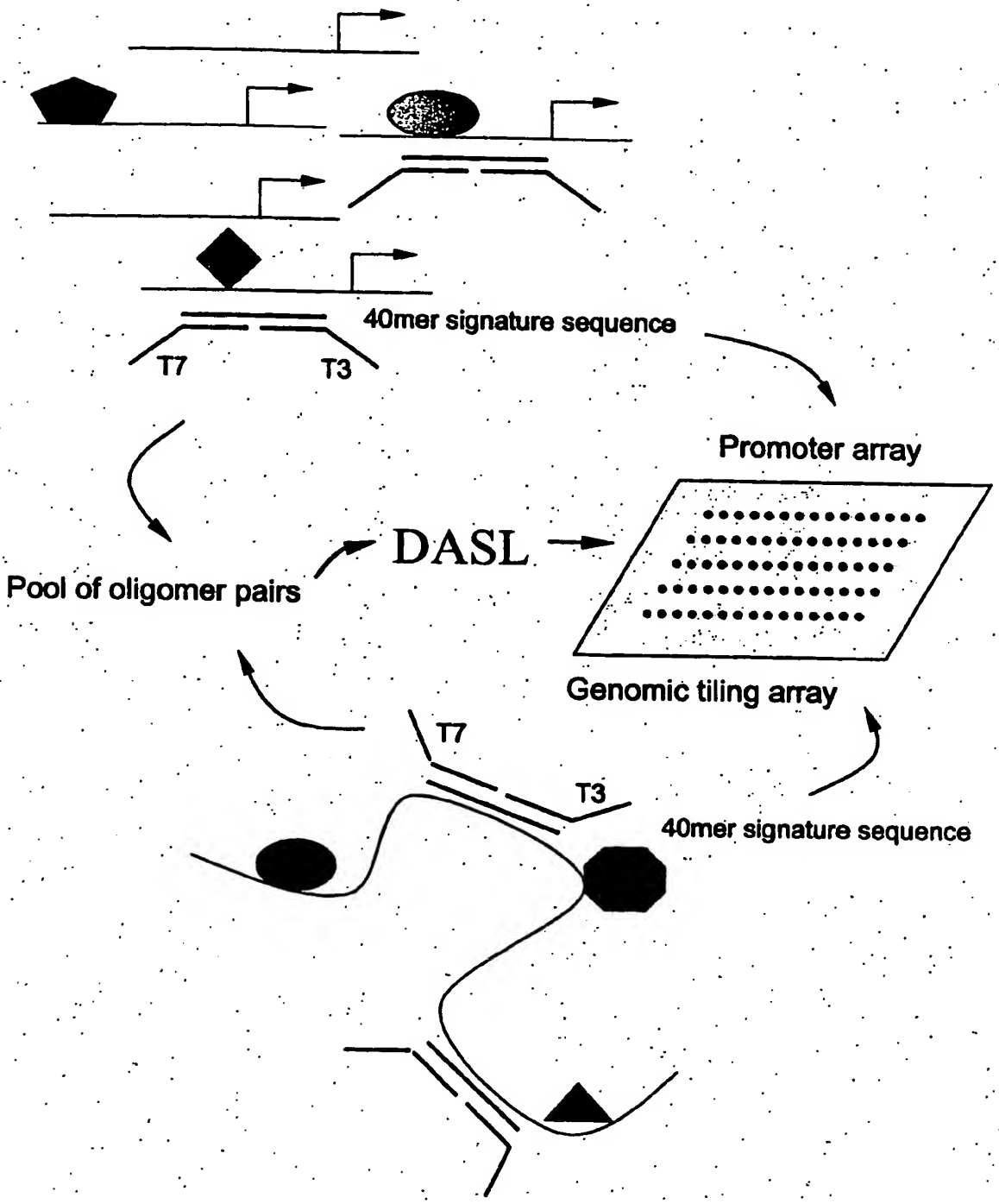
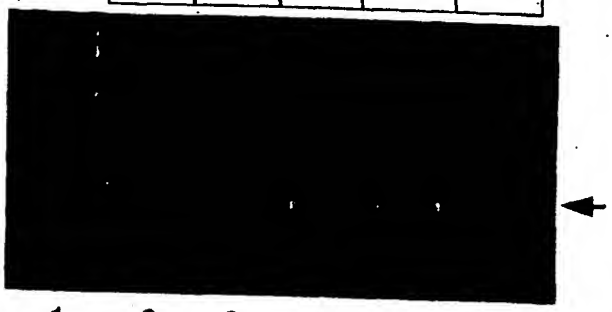


FIGURE 2

Specificity and Sensitivity of the ChIP-DASL Technology

A. Detection of specific genomic DNA sequences

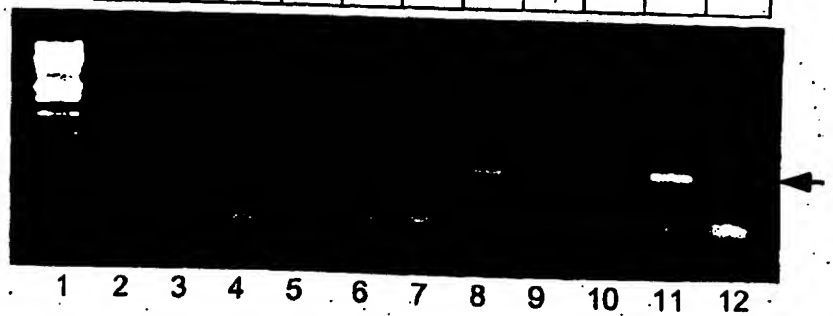
Genomic DNA		10^4	10^5	10^6	
Oligo Pool for Genomic DNA	+	+	+	+	



1 2 3 4 5 6

B. Detection of spiked sequences

Genomic DNA		10^5	10^5	10^5	10^5	10^5				10^5	
Spiked Plasmic DNA			10^2	10^3	10^4	10^5	10^5		10^5		
Oligo Pool for Genomic DNA								+	+	+	
Oligo Pool for Spiked DNA	+	+	+	+	+	+	+				



1 2 3 4 5 6 7 8 9 10 11 12

FIGURE 3

Detection of Spiked Human Promoters

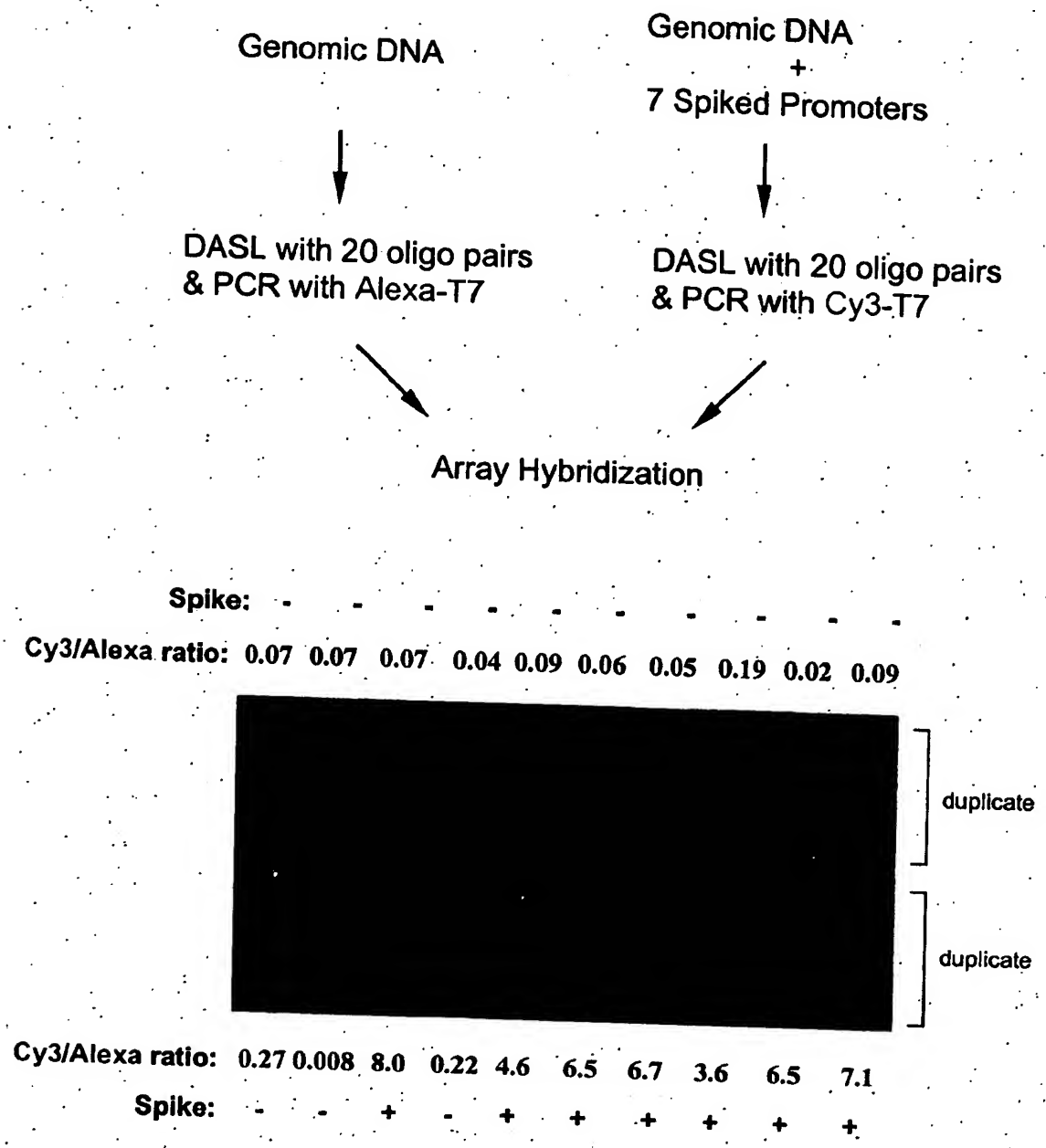
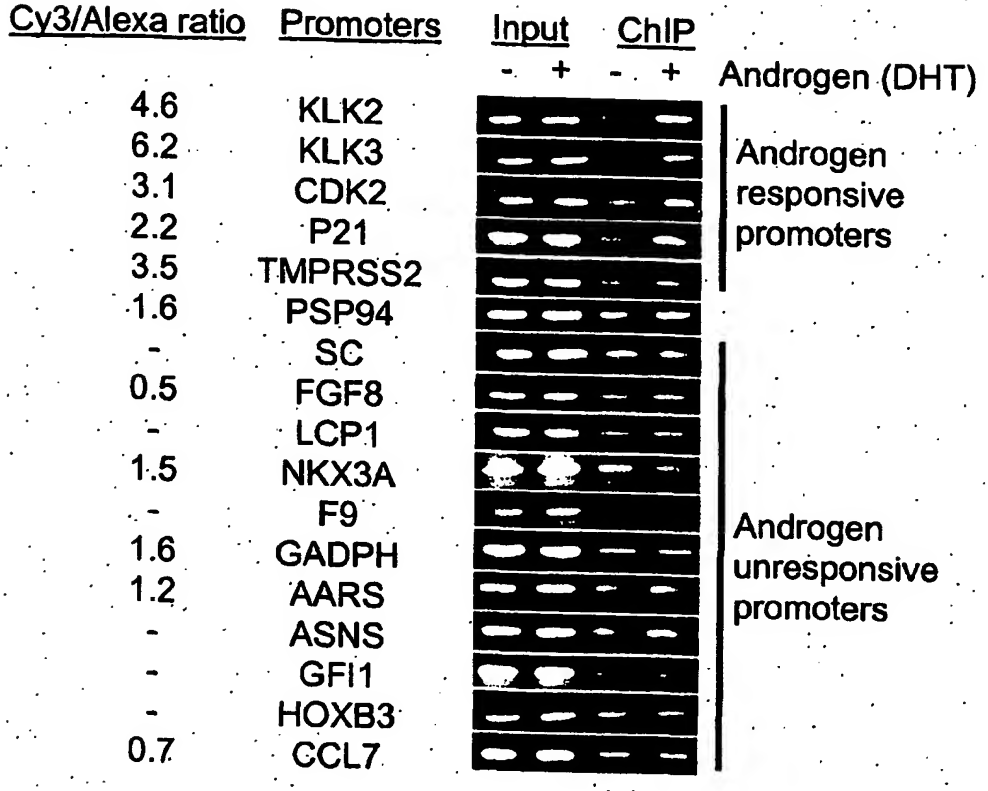


FIGURE 4

Confirmation of DASL results by conventional chromatin IP



Cy3 signal: ChIP with DHT-treated cells
Alexa signal: ChIP with mock-treated cells
"-": Background signal in both Cy3 and Alexa channels